

UniDive MC meeting 6

2 April 2025, 10:30-12:30 CEST, online

Agenda	1
Minutes	2
TODO	9
MWE 2025 workshop - selected papers	10

Agenda

1. Welcome to participants, verification of the quorum and adoption of agenda
2. Information to the MC
 - a. Recap of the minutes of the previous MC meeting
 - b. Action Membership
 - c. Grant Awarding by the Action
 - d. Budget status: summary from the Grant Holder
3. Follow up and discussion
 - a. Events in Grant Period 3
 - b. UniDive progress report at 24 months
 - c. Towards the Work and Budget Plan for Grant Period 4
 - d. Progress of each working group
 - e. Progress report from GPs 1-2
4. Work and Budget Plan for Grant Period 4
5. AOB
6. Next MC meeting

Minutes

Attendance:

Minuting: Agata Savary,

1. [AGATA SAVARY, ALINA WRÓBLEWSKA] Welcome to participants, verification of the quorum and adoption of the agenda
 - a. Fill in the [attendance list](#) (check a box under your name). Add your name in the row of the country you represent if needed.
 - b. The quorum is 2/3 of COST Full and Cooperating Member countries (one representative per country is enough)
 - c. Confirming the quorum: at the beginning of the meeting, 20 countries were represented but gradually this number increased; at the first MC vote 26 countries were represented; this number further increased and in total 27/38 countries were represented (71%)
 - d. MC vote rule: if two representatives of the same country cast different votes (approved vs. objected), the vote for this country is considered void
2. [AGATA SAVARY] E-votes and matters arising since the last meeting on 24 October 2024
 - a. Validating the MC5 minutes from 24.10.2024 - approved
 - b. Validating the 2 UniDive Progress Report - approved
 - c. Online vote to validate Nina HOSSEINI-KIVANANI (Luxembourg) as an MC member - approved
 - d. Decisions taken by the Core Group, on behalf of the MC - see past and upcoming [CG meetings](#)
 - e. Reminder: the MC delegated the following powers to the CG:
 - i. Approving/refusing WG applications - see below
 - ii. Selecting event participants who will be reimbursed - see below
 - iii. Adjusting the location and dates of activities (in close consultation with the Local Organizer) - see below
 - iv. Deciding about moderate budget expenditure (not exceeding 5000 EUR), to organize events which were not included in the Work and Budget Plan
 - f. No questions or comments from the MC
3. [DANIEL ZEMAN] Action Membership: *(as of March 16, 2025)*
 - a. New COST Members represented in the MC
 - i. 38 countries in total; new: Luxembourg
 - b. Action Participation: WG membership and applications
 - i. applications still coming almost every week, only 11 applications rejected since the action started, no new rejections in the last 11 months
 - ii. **404** people in total (members of at least one WG)
 - iii. 253 members of WG1
 - iv. 191 members of WG2
 - v. 256 members of WG3
 - vi. 163 members of WG4

- vii. 45 people are members of all four WGs
- viii. 1 MC member not in any WG
- ix. 7 former WG members terminated their membership
- x. Membership according to COST balance criteria:
 - 199 (49%) full member ITC, 168 (42%) full member non-ITC, 10 cooperating member (Israel), 3 near neighbor country (Kosovo, Tunisia), 12 international partner / organization, 12 without affiliation
 - 237 (59%) female, 160 (40%) male, 7 prefer not to say
 - 175 (43%) young researchers
- c. No questions or comments from the MC
- 4. [NURIT MELNIK] Grant Awarding by the Action, as of 12 March 2025 - year 3 (data added after the meeting)
 - a. Total budget: 55,000 EURO
 - b. Total granted: 32,676 EURO
 - c. 16 Grantees (out of 19):
 - i. 9 female; 6 male; 1 prefer not to say
 - ii. 7 young researchers; 9 not so young
 - iii. Affiliation: 5 ITC, 11 not ITC
 - d. Duration: 5-92 days (average 25 days)
 - e. Next call: Early May
 - f. Bug in the system: approving the applications on e-cost changes the dates and amounts of the STSMs. The bug was fixed.
 - g. Since 12 March one more STSM was granted
- 5. Grant Period 3 events
 - a. [VERGINICA MITITELU, VERONIKA LIPP] WG2 meeting and [3rd UniDive General Meeting](#) in **Budapest** - 28-30 January 2025
 - i. WG2 dedicated meeting: more than 65 participants, no parallel sessions -> possibility of a comprehensive perspective on the activities, common decisions wrt phenomena under investigations in several tasks, awareness of decisions with influence on tasks of interest, etc.; organized as round tables; online participation possible but with no interaction (not convinced that it is really helpful)
 - ii. Main meeting:
 - 122 participants
 - 3rd UniDive workshop with
 - Invited talk by Martin Haspelmath
 - 42 posters (peer-reviewed)
 - Extended Core Group meeting
 - WG sessions, no parallelism (appreciated)
 - iii. Selection of the event participants - all people who expressed their interest were invited (and reimbursed if eligible)
 - b. [JOAKIM NIVRE, KAJA DOBROVOLJC] Culminating workshop of the [1st UniDive Shared Task on Morphosyntactic Parsing](#) - **Ljubljana**

- i. Key Dates
 - March 7th: Training data release
 - July 1st: System submission deadline
 - August 26th: Half-day workshop in Ljubljana, Slovenia (co-located with [SyntaxFest 2025](#))
- ii. Reimbursement
 - Up to 40 participants will be reimbursed by COST (travel + 2 daily allowances)
 - Selection will be made in June by the Shared Task Organizers
- iii. Shared task description
 - Morphosyntactic parsing without word boundaries
 - Dependency tree over content words
 - Grammatical information as features
 - Data for six languages (more to come):
 - English
 - Italian
 - Polish
 - Portuguese
 - Serbian
 - Swedish
 - Evaluation:
 - LAS for dependencies
 - F1 score for features
 - Organizers:
 - Omer Goldman
 - Leonie Weissweiler
 - Reut Tsarfaty
- c. Two new WG events (not planned in the WBP3):
 - i. WG1 task 1.5 meeting on spoken data annotation - Ludovica Pannitto, Bologna - **29-30 May 2025**; see a [draft](#)
 - ii. WG1 meeting of the Turkic group, SyntaxFest - Soudabeh Eslami, Cagri Coltekin, Bermet Chontaeva, Furkan Akkurt - at SyntaxFest, 26th August 2 - see a [draft](#)
- d. No questions or comments from the MC
- 6. [BEATA WÓJTOWICZ] Budget status of year 3 - summary from the Grant Holder
 - a. Total budget EUR 227,000; we have already spent 38% - EUR 88,000 for the Budapest Meeting.
 - b. Reimbursement money for Budapest has been already sent.
 - c. Underspent Budapest budget: 34,910 EUR will contribute to STSMs, conferences and WG meetings. Proposed reallocation:
 - i. STSMs \Leftarrow 19,000 EUR (many STSM applications received)
 - ii. YRI and ITC conference grants \Leftarrow 5000 EUR
 - We propose to dedicate some grants to the MWE workshop 2025, colocated with NAACL 2025

- This workshop would be in the WBP 3 if it were collocated with ACL 2025 (Vienna)
 - The MC should decide if this is a high-level venue in the domain. This year the acceptance rate is 75%
 - Grants cover a fixed amount of money; attendance type: face to face or online
- iii. WG meetings \Leftarrow 10,000 EUR
- iv. Task 1.5 meeting on spoken data annotation \Rightarrow 5000 EUR
- v. Turkic group meeting \Leftarrow 5000 EUR
- d. Comments or questions from the MC
 - i. [KRISTINA] Is the MWE workshop the only possible venue for the conference grants or are other conferences also possible?
 - ii. [AGATA] Other conferences are also possible
- e. **Vote** about the re-allocation of the underspent budget as proposed above: 0 against, 0 abstaining, 26 in favor
- 7. [AGATA SAVARY] UniDive [progress report](#) at 24 months and the Rapporteur's review
 - a. Major issues highlighted by the Rapporteur
 - i. Insufficient dissemination to industry
 - ii. Delay of certain deliverables
 - iii. Not enough links with LLMs
 - iv. Other channels to use for communication
 - v. Moderate level of ambition of some objectives.
 - vi. Low level of ambition of some deliverables
 - vii. Criticism about the publications: Most publications are not peer-reviewed and published in self-organised venues
 - b. Our [request](#) for clarifications
 - i. COST response: the level-of-ambition items are automatically generated by the report interface and should not be seen as re-evaluation of what was already evaluated at proposal stage
 - ii. The Rapporteur agreed to have a call with the chair and the vice-chair to further discuss his suggestions
 - c. **Discussion:** How to take the review into account?
 - i. [LUCIA] Federation of multimedia can help to provide links with industry
 - ii. [DAVE] Possible industry links: Standardisation committees e.g. ISO/TC 37 Language & Terminology • Terminology Coordination Unit of the European Parliament, termcoord.eu • Federation Internationale des Traducteurs • EC Directorates-General for Translation and Interpreting • European Legal Interpreters & Translators Association (EULITA) • European Language Industry Association (ELIA)
 - iii. [TANJA] Has some links with industry via a project (UpSkills) - can provide links
 - iv. [TANJA] Use STSMs for visits in companies
 - v. [KAJA: added after the meeting] introduce a special theme or track for the next UniDive general meeting, dedicated to LLM-related work in the field

8. [AGATA SAVARY] Towards the Work and Budget Plan for Grant Period 4
 - a. 2nd training school & WG4
 - i. Proposals:
 - [Yerevan](#), Armenia (Anna Danielyan), resubmission from 2024
 - [Durrës](#), Albania (Manjola Zaçellari Lumani)
 - [Firat University in Elazığ](#), Turkey (Yelda Yesildal Eraydin)
 - ii. Selection criteria:
 - ITC country preferred
 - Involvement of the local team in the current UniDive tasks
 - Accessibility
 - iii. Core Group pre-selection: **Yerevan**
 - iv. Proposed dates: **January 19-24, 2026** (Monday-Saturday)
 - v. Daily allowance - proposed rates:
 - **185 EUR/day** for the trainers (following [COST rates](#))
 - **120 EUR/day** for the trainees (same as at the 1st training school in Chisinau, Moldova in 2024)
 - b. 4th general meeting
 - i. Proposals:
 - [Bucharest](#), Romania (Verginica Mititelu),
 - [Yerevan](#), Armenia (Anna Danielyan),
 - [Durrës](#), Albania (Manjola Zaçellari Lumani)
 - ii. Core Group pre-selection: **Bucharest**
 - iii. Proposed dates: **late August - early September, 2026**
 - iv. Daily allowance
 - Proposed rate: **130 EUR/day**
 - [VERGINICA] A 4-star hotel in the vicinity of the venue is about 80 euros/night. Of course we will try to negotiate some rates at some hotels as well.
 - c. No comments or questions from the MC
 - d. **Vote** about including the 2 proposed venues to WBP4: 0 against, 0 abstaining, **all** in favor
9. Progress of each working group
 - a. [BRUNO GUILLAUME, KAJA DOBROVOLJC] WG1 **Corpus Annotation**
 - i. Ongoing tasks ([refer to WG1 webpage for contacts](#)):
 - **T1.1. Linguistic typology and multilingual corpus annotation**
 - Survey about "Annotation practices in Universal Dependencies".
 - 65 participants
 - Main interests: linguistic research: 46%, NLP: 29%, annotation: 18%, Digital Humanities, Translation studies
 - See more in these [slides \(13 to 22\)](#)
 - **T1.2. MWE annotation guidelines and UD-PARSEME unification**
 - [PARSEME 2.0 annotation campaign](#) started (kick-off meeting 20 March)
 - Annotation of all MWE (only verbal in previous editions)
 - 10 to 15 languages
 - April-May: Annotation
 - 15 July: Sample data for a SemEval task

- **T1.3. Extensions and updates to morphosyntactic annotation guidelines**
 - **Subtask A: Discussion on the current guidelines:**
 - Use of GitHub issues in the [UD Issue Tracker](#).
 - UniDivers are invited to create new items there if they have issues that are not yet present
 - **Subtask B: Construction-oriented guidelines.**
 - In addition to the current UD guidelines with pages for part-of-speech tags, morphological features and relations in individual languages, a new incubator for [construction-oriented documentation](#) is currently built. The goal is to provide a larger picture with a wholistic solution for individual constructions and strategies with a typologically varied perspective.
- **T1.4. Sharing tools, formats, and infrastructure**
 - **Subtask Manual annotation tools**
 - Overview of available tools for manual annotation (more specifically on morpho-syntax and MWE).
 - 21 tools covered
 - A [interactive webpage](#) to explore the tools
 - Finalisation and publication planned
 - More in these [slides \(25 to 35\)](#)
 - **Subtask File Format**
 - Agreement on limitations of tabular formats
 - An alternative more structures format (based on JSON or XML) will be proposed
 - See [slides from Budapest](#)
- **T1.5. Annotation of Spoken data**
 - In UD, 40 treebanks with spoken data in v2.14 (36 languages and 22 languages family)
 - Survey sent to treebanks developers: 22 treebanks responded.
 - Very different type of spoken data
 - Tendency toward verbatim transcription with rules of standard orthography
 - Metadata is missing, unharmonized, hard to discover
 - Annotation guidelines for some speech-specific phenomena are underspecified or poorly documented
 - More in these [slides \(4 to 12\)](#)
 - A dedicated workshop is planned in May 2025
- ii. Next meeting: Tuesday 15 April 2024 from 14:00 to 15:30 CEST/Paris
- iii. No comments or questions from the MC
- b. [VERGINICA MITITELU, VOULA GIOULI] WG2 **Lexicon-corpus interface**
 - i. Ongoing tasks:
 - **T2.1.1. Harmonizing the definition of a “syntactic word” across languages** - The survey based on Martin Haspelmath's paper on [Word](#), filled in by 26 participants, is being discussed in meetings dedicated to language families. Afterwards, conclusions will be included in a paper.
 - **T2.1.2. Harmonizing lemmatization rules (for words and MWEs) and lexical features across languages** - A working package was prepared for volunteers to fill in regard to lemmatization practices in UD treebanks

for single words. At the moment the information was filled for several UD treebanks.

- **T2.2. Design of a lexicon-corpus interface** - extension of the ELEXIS-WSD Parallel Sense-Annotated Corpus with new languages: still in the process of checking translations (Macedonian, Greek), finished checking the automatic translations (Croatian, Serbian, Polish), checking POS annotation (Romanian). Of the existing 10 languages already included in the corpus, the Dutch team is correcting automatic UD parsing; the Estonian team has recently finished and will proceed with MWE identification in the corpus; providing sense inventories for corpus annotation for Croatian (conversion of the Croatian School Dictionary), Ukrainian (Ukrainian WordNet + Boris Grynchenko Dictionary of Ukrainian), and Serbian (Serbian WordNet)
 - **T2.3.1. Survey on MWEs lexicons** - weekly meetings - final polishing - chapter to be submitted to an edited volume (LSP MWE series)
 - **T2.3.2. Survey on MWE lexica-aware NLP tasks/applications** - bi-weekly meetings - almost finished, 2 sections (out of 7) still requiring revisions - to be submitted as a chapter to an edited volume (LSP MWE series)
 - **T2.4. Canonical forms of MWEs** - a draft working document on canonical form for MWEs; examples from several languages necessary.
 - **T2.5. Extensions of UD-PARSEME annotation for constructions and constructicons** - a survey was sent out and responses are still waited for
 - **T2.3.3. Standardizing lexicons of MWEs** - to set off
- ii. Next meeting: 30 April
- iii. No comments or questions from the MC
- c. [JOAKIM NIVRE, GÜLSEN ERYIGIT] WG3 **Multilingual And Cross-Lingual Language Technology**
- i. Ongoing tasks:
- **Task 3.1 Documentation of multilingual tools and resources**
 - Leaders / Contacts: A. Seza Doğruöz, Teresa Lynn, Maria Giagkou
 - Status: First phase completed. Planned: Documentation of systems participating in evaluation campaigns.
 - **Task 3.2 Evaluation campaign: morphosyntactic parsing**
 - Leaders / Contacts: Omer Goldman, Leonie Weissweiler, Reut Tsarfaty
 - Status: Training data released for six languages, more languages to come. Release of test data 15 June. System submission 1 July.
 - **Task 3.3 Conceptions of multilinguality**
 - Leaders / Contacts: Adriana Pagano, Ilan Kernerman
 - Status: Survey completed. Report in progress.

- **Task 3.4 Evaluation campaign: PARSEME 2.0: A Multilingual Shared Task on Identification, Generation and Understanding of Multiword Expressions**
 - Leaders / Contacts: Manon Scholivet, Agata Savary, Takuya Nakamura and Carlos Ramisch
 - Status: Proposal submitted to SemEval 2026.
- **Task 3.5 Evaluation campaign: AdMIRe - Advancing Multimodal Idiomaticity Representation**
 - Leaders / Contacts: Thomas Pickard , Aline Villavicencio
 - Status: First edition at SemEval 2025. Second edition in the planning phase.
- ii. Next meeting: May 6, 12:00 CEST
- iii. No comments or questions from the MC
- d. [MARIE-CATHERINE DE MARNEFFE, OLHA KANISHCHEVA] WG4
Quantifying and promoting diversity
 - i. **Ongoing tasks:**
 - Task 4.1:** Promoting low-resourced/endangered languages [co-leaders: Lucia Amoros, Abigail Walsh]
 - Task 4.2:** Survey of diversity measures [leader: Louis Estève]
 - Task 4.3:** Measuring diversity of shared tasks [co-leaders: Esther Ploeger and Wessel Poelman]
 - Task 4.4:** Benchmarking for diversity [co-leaders: Manon Scholivet and Alina Wróblewska]
 - ii. Next meeting: To be decided after the call for trainers at the winter school goes out; dedicated WG4 in-person meeting co-located with the winter school in January 2026
 - iii. No comments or questions from the MC
- 10. Next MC meeting
 - a. **Thursday 23rd October 2025 10:00-12:00 online**
- 11. Summary of MC decisions
 - a. Reallocating the saving from Budapest to STSMs, conference grants and 2 ad-hoc WG1 meetings
 - b. Validating Yerevan and Bucharest as venues for the 2nd training school and final general meeting

TODO

- ~~[AGATA] Publish these minutes and submit them to an MC vote.~~
- ~~Future events:~~
 - ~~[AGATA] Create webpages: Bologna, Ljubljana – Turkic workshop, Yerevan – training school, Yerevan WG4~~
 - ~~[AGATA] Inform the event proposers: Ludovica, Soudabeh, Anna, Verginica~~
 - ~~Vote the daily allowance for Bologna~~
- ~~[NURIT, STELLA] Consider STSM in companies in the next call → see the CG 25 agenda~~

- ~~[VERGINICA] Share the list of the papers selected for the MWE workshop → see below~~
- [AGATA] Newsletter with future events

MWE 2025 workshop - selected papers

- *Syntagmatic Productivity of MWEs in Scientific English*
Diego Alves, Stefan Fischer and Elke Teich
- *Probing Internal Representations of Multi-Word Verbs in Large Language Models*
Hassane Kissane, Achim Schilling and Patrick Krauss
- *VMWE identification with models trained on GUD (a UDv.2 treebank of Standard Modern Greek)*
Stella Markantonatou, Vivian Stamou, Stavros Bompolas, Katerina Anastasopoulou, Irianna Linardaki Vasileiadi, Konstantinos Diamantopoulos, Yannis Kazos and Antonios Anastasopoulos
- *Using LLMs to Advance Idiom Corpus Construction*
Dogukan Arslan, Hüseyin Anıl Çakmak, Gulsen Eryigit and Joakim Nivre
- *Gathering Compositionality Ratings of Ambiguous Noun-Adjective Multiword Expressions in Galician*
Laura Castro and Marcos Garcia
- *Survey on Lexical Resources Focused on Multiword Expressions for the Purposes of NLP*
Verginica Mititelu, Voula Giouli, Gražina Korvel, Chaya Liebeskind, Irina Lobzhanidze, Rusudan Makhachashvili, Stella Markantonatou, Aleksandra Markovic and Ivelina Stoyanova
- *A European Portuguese corpus annotated for verbal idioms*
David Antunes, Jorge Baptista and Nuno J. Mamede
- *MultiCoPIE: A Multilingual Corpus of Potentially Idiomatic Expressions for Cross-lingual PIE Disambiguation*
Uliana Sentsova, Debora Ciminari, Josef Van Genabith and Cristina España-Bonet
- *Named Entity Recognition for the Irish Language*
Jane Adkins, Hugo Collins, Joachim Wagner, Abigail Walsh and Brian Davis